AI Policy

We are leveraging AI to enhance our products, focusing on scenarios where there is minimal or no risk of hallucinations and ensuring graceful handling of errors. This allows us to maintain control over potential issues and resolve them effectively.

Guiding Principles

Our AI use cases include, but are not limited to:

- Company summarisation
- Keyword extraction
- Sector classification
- Company comparison
- Company similarity
- News matching
- Company disambiguation
- Named entity recognition (names, countries, company names, etc.)

Models

We utilise a range of technologies tailored to each use case, from traditional ML models like Random Forest classification to advanced LLMs such as Llama 3 and GPT-4.

Our team has extensive experience in building and training bespoke models, with a strong emphasis on quality and thorough evaluation. We understand that LLMs can exhibit variability in responses, and we measure model performance quantitatively, considering this random behaviour. Human-annotated data is crucial for achieving the best results, as if we input bad data, the output will not give us the desired results.

Framework and Service Providers

Depending on the request and use case, we either run models ourselves or rely on third-party services. When building frameworks and models in-house, we use standard open-source tools like scikit-learn, Huggingface Transformers, and PyTorch. Our mature MLOps capability ensures model versioning and secure storage of training data, all within our hybrid cloud infrastructure.

Third-Party Models and APIs

When appropriate, we use third-party model APIs such as OpenAI GPT-4, Anthropic Claude 3, and Google Gemini. We always ensure we have your consent before using these services with sensitive data and utilise enterprise versions that comply with geographic hosting commitments and data privacy regulations. Our external LLMs are accessed through a secure internal gateway, ensuring full awareness of data sharing.

For those who prefer not to use third-party services, we offer internal LLMs hosted securely in our hybrid cloud, with all data encrypted in transit.

Data Management

External Models

When using external models, we log requests securely for diagnostic purposes. Detailed logs, including prompts and responses, help us trace and address any hallucinations.

Internal Models

Your data remains proprietary, and we do not use it for training without explicit permission. We log traffic for diagnostic purposes but do not retain detailed input and output data. Final results are stored securely in the Syfter database, isolated from other instances.

Data Handling and Protection

Automated decision-making processes at Syfter involve analysing and enriching company profiles. Most automated models are internal, and we do not send your data to third-party models without prior discussion and approval.

All communication is centrally managed and encrypted, with individual client requests stored separately.

Logs are retained only for a specified duration or a maximum of 30 days.

Our hybrid cloud, combining GCP and Filament-owned servers, ensures data security.